

<https://doi.org/10.22363/2687-0088-26828>

Research article

A very unpredictable ‘person’: A corpus-based approach to suppletion in West Polesian

Kristian RONCERO  

Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
Friedrich Schiller Universität, Jena, Germany
 kristianroncero@gmail.com

Abstract

In Slavic languages, as in many other languages, the noun for ‘person’ has a suppletive paradigm. Yet, as this study shows, in West Polesian (East Slavic) the noun ‘person’ is a typological outlier not only within Slavic but also cross-linguistically because it combines three stems with a very complex distribution. This paper looks for any regularities in the distribution of these suppletive stems, their cognates among other Slavic languages and how speakers use them in free texts. This survey provides novel insights into suppletion. First, suppletion involving more than two stems is typologically uncommon but the West Polesian noun ‘person’ combines three. Second, against any expectation of regularity for the sake of learnability, free-text data show that speakers do not distribute the stems homogeneously. Third, notwithstanding the diglossic situation in Western Polesie, the inter- and intra-speaker variation in the choice of stem does not seem particularly conditioned by sociolinguistic variables such as gender, age or social class. In sum, this corpus survey of the suppletive stems of ‘person’ in West Polesian and Slavic illustrates a rare case in morphological typology where there is a three-stem suppletion combined with overabundance and a vast amount of variation across speakers.

Keywords: *corpus, field-data, overabundance, Slavic, suppletion, variation, West Polesian*

For citation:

Roncero, Kristian. 2022. A very unpredictable ‘person’: A corpus-based approach to suppletion in West Polesian. *Russian Journal of Linguistics* 26 (1). 116–138
<https://doi.org/10.22363/2687-0088-26828>

Непредсказуемый «человек»: корпусное исследование супплетивности в западнополесских говорах

Кристиан РОНСЕРО  

Институт эволюционной антропологии им. Макса Планка, Лейпциг, Германия

Университет Фридриха Шиллера, Йена, Германия

 kristianroncero@gmail.com

Аннотация

В славянских языках, как и во многих других, существительное, обозначающее «человек», имеет супплетивную парадигму. Однако, как показывает данное исследование, в западнополесских (восточно-славянских) говорах существительное «человек» типологически аномально не только для славянских, но для других языков, так как представляет собой комбинацию трех основ с очень сложной дистрибуцией. В статье рассматривается дистрибуция этих супплетивных основ, родственные им слова в других славянских языках и их использование носителями в произвольных текстах. Это исследование предлагает новый взгляд на супплетивность. Во-первых, супплетивность, включающая более двух основ, нетипична, но в западнополесском существительном «человек», используются три основы. Во-вторых, вопреки ожиданиям, данные произвольных текстов показывают, что говорящие не распределяют основы равномерно. В-третьих, несмотря на ситуацию диглоссии в Западном Полесье, внутри- и межличностные вариации в выборе основ не особо зависят от социолингвистических переменных, таких как гендер, возраст или социальный класс. В целом, это корпусное исследование супплетивных основ существительного, обозначающего человека в западнополесском и славянских языках, иллюстрирует редкий случай морфологической типологии, где присутствует трехосновная супплетивность вкупе с избыточностью и высоким уровнем вариативности среди носителей языка.

Ключевые слова: корпус, данные полевых исследований, избыточность, славянские языки, супплетивность, вариация, западнополесские говоры

Для цитирования:

Roncero K. A very unpredictable ‘person’: A corpus-based approach to suppletion in West Polesian. *Russian Journal of Linguistics*. 2022. Vol. 26. № 1. P. 116–138 <https://doi.org/10.22363/2687-0088-26828>

1. Introduction

Bybee (1985: 91) defined suppletion or suppletive paradigms as “inflectional paradigms that have forms built on two or more stems that are etymologically from different sources.” I give more detail on suppletion later on, but as a more familiar example we can think of the English verb *to go* in the PAST SIMPLE > *went* (**go-ed*; as in *jump* > *jump-ed*). Both stems are phonologically very different, furthermore, there is evidence that etymologically they derive from two different roots (OED 2018). Hence, *go* and *went* hold a suppletive relation; that is to say, their correlation is semantic, rather than formal (phonological).

West Polesian is a little-known East Slavic variety spoken between southwestern Belarus, north-western Ukraine and a small fraction of eastern Poland. The

speech community lives in an area with a swampy topography that is frequently flooded in spring. As a result, speakers have been highly isolated from outside groups for centuries, as well as having heavily limited mobility within this region. The isolation has contributed to the preservation of some older Slavic cultural and linguistic features, whilst it has also set the ground for innovations with respect to the East Slavic family. Nowadays the community is increasingly exposed to the surrounding standardized and closely related Slavic varieties (i.e. Belarusian, Russian, Ukrainian and Polish). The language contact resulting from media and education is putting pressure on West Polesian grammar. Nevertheless, this is not the only source of contact and pressure. Since the 1980s the government has been draining the marshes, partly, in order to build roads. This has led to a massive emigration of the younger population to the cities, where, in the case of Belarus, Russian is spoken. As a result, West Polesian grammar can often feel like a crossroad of four main Slavic varieties.

Once I advanced on the transcription of the recordings of the West Polesian corpus, I realized that the noun ‘person’ displayed a peculiar behavior, which differed from what I knew from its Belarusian, Russian or Polish cognates. I was aware of the alternation between the stems *ʃolovik-* and *lud-*, in West Polesian. Yet, further on, I noticed that a third stem, *duf-*, was another juggling ball of the suppletive paradigm of the noun ‘person’:

- (1) (B20.17 00:25)

i	jak	ʃolo'vik	ide	noʃu	obizatelno	puzaj-e
and	when	person.NOM.SG	go-3SG	at night	necessarily	scare-3SG

‘And when a person/man (= someone) walks [over] he always scares them.’
- (2) (B20.19 01:51)

teper	uže	tak-ix	ludej	praktičeski,
now	already	these-GEN.PL	person.GEN.PL	virtually
poʃt	i	ni-ma		
almost	and	NEG-HAVE		

‘Nowadays there are hardly any people like this left [sorcerers].’
- (3) (T1.18 01:16)

[...]	bo	pjat,	sjem	duf...	na	sjem	ʃolo'vik...
	as	five	seven	person.GRADNM	to	seven	person.GRADNM

‘[...] because [there were] five, seven people... for seven people.’

At first glance, it is tempting to assume that these forms are independent lexical entries (synonyms), as traditionally OVERABUNDANCE (Thornton 2019) has been despised. However, a further cross-Slavic corpus survey showed that *duf-* had cognates in other Slavic languages as a suppletive stem of ‘person’ (see the cross-Slavic survey in (§4.1.)). Having a three-stem inventory, (instead of two) like the vast majority of Slavic languages, would make this phenomenon exciting enough to be studied. But there is an even more interesting twist; despite the inventory of

suppletive stems being larger than normal their distribution is very heterogeneous across speakers (and even within their own idiolect) when comparing their use in free texts. That is to say, suppletion makes morphology more complex to learn and retrieve (Bybee 1985), so the more stems involved the more regularity we would expect to compensate it, but I will refute this hypothesis in this article.

Numeral Phrases (NumPs) and government are closely related to the stem alternations of the nouns ‘year’ and ‘person’ in Slavic. For this reason, I start with a short overview of their morphosyntax and outline the particularities of Slavic and specifically West Polesian (where there are dedicated ADNUMERATIVE forms) (§2) in order to frame this study. Second, I introduce the study, present the methodology (§3) and the results from a cross-Slavic survey of the noun ‘person’ in combination with NumPs (§4.1). This shows that the three stems discussed here have cognates in other Slavic varieties. Third, I present some ‘ideal paradigms’ of the three stems involved in West Polesian (§4.2), admitting that these paradigms are frequently combined and mixed by the speakers. Furthermore, the stems *folo’vik-* and *duf-* present complications for the analysis, as homophonous forms exist with full paradigms (§4.2.1). Fourth, based on Bortnik (1979) and Chumakina et al. (2004) I propose some putative conditions in order to prove whether the choice of one stem over another is restricted and/or motivated by these (§4.3). Moreover, I prove that sociological factors are not automatic predictors of the use of one form or the other, either (§4.3.3). Finally, I present a summary and conclusions extracted in the light of the inconsistencies between speakers and the peculiar sociolinguistic setting of Western Polesie (§5).

2. Some remarks on Numeral Phrases in Slavic and the peculiarities of WP

Most of the suppletion in the paradigm of ‘person’ in West Polesian and Slavic happens around the cells used with Numeral Phrases (NumPs) and quantification. The syntax of NumPs and quantification is very complex in Slavic. The topic has been widely discussed in the literature (and is still being debated). For more complete descriptions and hypotheses from colleagues belonging to different schools I shall refer to the following: Akiner 1983, Babby 1987, Corbett 1983, Franks 1995, Kim 2009, Madariaga & Igartua 2017, Mel’čuk 1985, Nessel 2019, Nessel & Nordrum 2019, Pereltsvaig 2013, Viellard 2011 and Žolobov 2003, among others. Hence, I do not intend to add anything to this topic, but rather I shall mention some of the peculiarities of West Polesian.

Common Slavic had a DUAL NUMBER that has been lost or heavily eroded in most contemporary Slavic languages with the exceptions of Slovene and Upper and Lower Sorbian. In Common Slavic the numeral ‘one’ governed SINGULAR; ‘two’, DUAL; ‘three’ and ‘four’ NOM PL; and higher numerals GEN PL (Akiner 1983, Žolobov 2003). Due to some phonological and morphosyntactic changes, the numerals ‘two’, ‘three’ and ‘four’ (henceforth,

LOWER NUMERALS) ended up merging in most Slavic languages (concerning morphosyntax). West Polesian has developed a dedicated form for nouns governed by LOWER NUMERALS called ADNUMERATIVE (ADNM) or NUMERATIVE (Mel'čuk 1985, Nurmio & Willis 2017, Roncero 2021, Žolobov 2003). In West Polesian, this ADNUMERATIVE (henceforth, ADNM) form is often in competition with what (at least phonologically) resembles NOM PL and GEN SG (4).

- (4) (B6, B9.elicited)
- | | | |
|-------------------|-----------|-----------------|
| a. ^I | dv-a | 'duba |
| | two-NOM.M | oak(M).GEN.SG |
| a. ^{II} | dv-a | du'bi |
| | two-NOM.M | oak(M).NOM.PL |
| a. ^{III} | dv-a | 'dubi |
| | two-NOM.M | oak(M).ADNM |
| | | 'Two oak trees' |

Moreover, at least some West Polesian nouns seem to also have a special noun form when governed by a HIGHER NUMERAL (i.e. different from the regular GEN PL), which I call GREATER ADNUMERATIVE (GRADNM) (5).¹

- (5) [elicited]
- | | | | |
|------------------|-------------|-----------------------|------------------|
| a. ^I | pjetj | lu'dej | |
| | five.NOM.PL | person.GEN.PL | |
| a. ^{II} | pjetj | tfolo'vik | |
| | five.NOM.PL | person.GRADNM | |
| | | 'Five people' | |
| b. ^I | saraj | fiet-ix | lu'dej |
| | barn.NOM.SG | DEM-GEN.PL | person.GEN.PL |
| b. ^{II} | *saraj | fiet-ix | tfolo'vik |
| | barn.NOM.SG | DEM-GEN.PL | person.GRADNM |
| | | 'These people's barn' | |

Even if forms like *tfolo'vik* were only GEN.PL, the solution is to propose that another 'non-canonical phenomenon' known as HETEROCLISIS is behind this (Stump 2006). Heterocclisis consists of using two inflectional forms belonging to two different inflectional classes. In this case, the *tfolo'vik* (and similar forms for 'year' e.g. *pjet hod*, instead of the regular *fiu'dru* 'five years', which is also attested) would have a GEN.PL proper of inflectional class I,² unlike the SINGULAR sub-paradigm, which belongs to class II. Note that in West Polesian (as in East Slavic) the nominal paradigms of classes I and II are only distinct from each on the ACC/GEN cell(s) of the PLURAL sub-paradigm.

¹ In Mel'čuk's (1985: 430–437) terminology this is an "adnumerative plural" but explaining the choice of my term would involve a long discussion deserving, at least, a paper on its own to provide enough context.

² Traditionally, in Slavic descriptions this is generally referred to, though inaccurately, as feminine (PACE Corbett).

3. Methodology and some representation remarks

In order to make sure that we are not dealing with overlapping synonyms, I have undertaken a cross-Slavic survey to see the behavior of the noun ‘person’ (Table 1; Table 2; Table 3). The materials for the cross-Slavic survey are based on *ParaSol, a Corpus of Slavic and Other Languages* (Waldenfels & Meyer 2011).

The corpus from which the West Polesian examples in this article have been extracted is the result of eight months of my own fieldwork in the region of Brest (Belarus).³ More than fifty speakers took part in the project, contributing different types of data. In order to protect the anonymity of the participants who did not want to disclose their identities each was assigned a code. The first two/three letters of the speaker code indicate the village where they come from. I will refer to this information further on for comparison. It should be noted that all the examples used in this paper have been exclusively taken from free texts, which have given 245 tokens for ‘person’. For the sake of transparency and data replicability, the full list of tokens can be found in the *Supplementary Materials*.⁴

The examples in this paper are transcribed according to IPA conventions (except for capitals in proper names), based on my ongoing (impressionistic) analysis of West Polesian phonology. I have tried to respect all the differences in the pronunciation, which means that sometimes stems appear as palatalized (*fʲelovjek-*, *ʲud-*, as in Contemporary Standard Russian – CSR); or most often unpalatalized (which is more in line with the general phonological rules of West Polesian) and the vocalism is not very consistent.

West Polesian stress is dynamic and has a lexical function. In West Polesian many nominal (and less so verbal) paradigms have mobile stress which helps to disambiguate otherwise segmentally syncretic forms (much more than in Belarusian and Russian); e.g. (B6) [GEN SG] *ka 'navɪ* vs. [NOM PL] *kana 'vɪ* ‘channel(s)”; (Z4) [GEN SG] *'pɪsni* vs. [NOM PL] *pɪs 'ni* ‘song(s)’. Thus, given that in West Polesian CASE/NUMBER marking is more clearly dependent in both stress and suffixation than in other Slavic languages, I do not provide morphological segmentation for nouns. Whenever there are differences regarding the position of the stress between speakers or even the same idiolect I have respected these. Unless specifically glossed, all numerals are cardinals.

4. The noun ‘person’

Under this heading, firstly, (§4.1) I present the noun ‘person’ and its behavior with numeral phrases across the Slavic family distinguishing synonyms from pure suppletive stems. Secondly, I present West Polesian paradigms for ‘person’ to which in an *ideal* instance speakers would adhere (§4.2); also mentioning some of the interferences derived from homophony

³ See §Appendix I for more details on the villages covered.

⁴ See list of tokens of ‘person’ on the West Polesian corpus at <https://doi.org/10.5281/zenodo.5879004>

(§4.2.1). Thirdly, I propose several putative conditions which could affect the choice of one stem or another, starting with morphosyntax (§4.3.1) and (§4.3.2), and moving to sociolinguistics (§4.3.3).

4.1. The noun ‘person’ across the Slavic family

I ran a cross-Slavic survey of the noun ‘person’ using the corpus by Waldenfels & Meyer (2011). Visualizing the suppletive stems and the available distributions for these across the Slavic family has been enlightening for this study. The cells most affected by suppletion are the ones that are most often employed for numerals (i.e. ADNUMERATIVE (ADNM), GENITIVE PLURAL/GREATER ADNUMERATIVE (GRADNM) or similar). That is why, in the following table, I pay special attention to the type of numeral each form appears with in DIRECT CASES.

Here is a list of the languages included in the survey, classified according to their subfamilies:

South Slavic⁵: Bulgarian (BG); Croatian (HR); Macedonian (MKD); Serbian (SRB); Slovene (SVO).

West Slavic: Czech (CZ); Polish (POL); Slovak (SK); Upper Sorbian (US).

East Slavic: (Standard) Belarusian (BLM); Contemporary Standard Russian (CSR); (Standard) Ukrainian (ULM).

Since the results of this survey are quite heterogeneous within each sub-family, I present them according to their genetic/areal affiliation rather than sorting them by the stems. Where there have been many forms, I have stressed in bold the most common or dominant form. The areas in grey indicate the absence of data or results.⁶

Table 1. South Slavic

	‘nom sg’	lower numerals		higher numerals	‘nom pl’
BG	čovek	čoveka		čoveka	čoveci
					xora
		duši		duši	duši(te) ⁷
HR	čovjek	čovjeka		ljudi	ljudi
SVO	človek	[2] človeka	[3] ljudje	ljudi	ljudje
MKD	čovek	čoveka ⁸		duši	duši(te) ⁹
		duši			
		luđe		luđe	luđe
SRB	čovek	čoveka		Ljudi	ljudi

⁵ In this survey, I distinguish Macedonian (MKD) from Bulgarian (BG) and Serbian (SRB) from Croatian (HR) as I obtained significantly different results for these pairs; however, I did not find any corpus also covering Bosnian and Montenegrin.

⁶ Although the vast majority of results come from observations from the *ParaSol* corpus (Waldenfels & Meyer 2011), I had to use *Hrvatski Jeyzčni Portal* (2006–2021) and *Rečnik na bālgarskija ezik* (2018) in order to confirm some of the results, which were otherwise inconclusive by mere observations on the corpus.

⁷ This form is very marginal and all the corpus results point out that, unless an article is used, the stem *duši* can only be used with quantifiers. Moreover the *Dictionary of the Bulgarian Academy (Rečnik na bālgarskia ezik, 2018* (online) also notes that *duši* must be used with quantifiers.

⁸ Only one hit in the entire corpus.

⁹ As in Bulgarian, this form is marginal and can only appear with an article.

Table 2. West Slavic

	‘nom sg’	lower numerals	higher numerals	‘nom pl’
CZ	člověk	lidé	lidí	lidé
POL	człowiek	ludzie / ludzi	ludzi	ludzie
SK	človek	ľudia	ľudí	ľudia
US	čłowjek	[2][n.d.]	[3-4] ludźo	ludźo

Table 3. East Slavic

	‘nom sg’	lower numerals	higher numerals	‘nom pl’
BLM	čalavek	čalaveki	čalavek ljudzej	ljudzi
CSR	čelovek	čeloveka	čelovek ljudej	ljudi
ULM	ljudyna	ljudyny	čolovik ljudej	ljudy

Due to time limitations, I focus only on suppletion in contexts where there are numerals; i.e. where dedicated ADNUMERATIVE forms are used (cf. Roncero 2021),¹⁰ and NOM SG – NOM PL, in order to see the contrast.

Whilst the noun ‘person’ in West and East Slavic is fairly unexciting, there is a very rich variation in the South Slavic branch. Other synonyms have been interfering with corpus results. The forms *osoba* (and the like) and *lica* (and the like) are present in many Slavic languages alongside the more established forms (*človek*, *ljudi*, *duši* and the like). In most languages I have found enough evidence for stating that such forms are not additional stems, but just synonyms. For example, the following sentence would not be allowed in Slovak if the stems were suppletive (i.e. two suppletive stems in a disjunctive) “<...> obyčajní ľudia či osoby nevedomé” ‘<...> ordinary people or unknown persons’ (Waldenfels & Meyer 2011: 77–439)). After refining the results of the corpus, we can see that the suppletive stems under discussion (for West Polesian ‘person’) are present in other Slavic languages, particularly in the South Slavic branch. However, it is important to remark that the stem *duf-* does not appear elsewhere in (contemporary) East or West Slavic, so this is probably an archaic feature of West Polesian, rather than a contact-induced loan.

4.2. West Polesian suppletion patterns for ‘person’

Based on observations from the corpus, these are the underlying paradigms of the three stems for ‘person’. Note that for *lud-* many speakers produce it as [lud-].¹¹ Nonetheless, for the sake of consistency (and to focus exclusively on suppletion), I have kept the non-palatalized form, in the paradigm below (Tables 4 a, b, c). Thus, I represent the cells containing those stems following a color code: *řolovik-*, red; *lud-*, blue; *duf-*, yellow. Given that the status of the GREATER ADNUMERATIVE is arguable, I have represented such forms as GEN.PL.

¹⁰ Regardless of the morphosyntactic strategy they follow, some nouns in ADNUMERATIVE may have a morphophonologically dedicated form, others may use the regular NOM PL form.

¹¹ And for most speakers there is free variation in the non-direct cases between the two realizations.

Tables 4 a, b, c Available stems for the noun ‘person’ in West Polesian

(a) the stem *ʃolovʲk-*

	SG	PL
NOM	ʃolo'vʲk	
VOC	ʃolo'vʲʃe!	
ACC	ʃolo'vʲka	
GEN	ʃolo'vʲka	ʃolo'vʲk
DAT	ʃolo'vʲku	
INST	ʃolo'vʲkom	
LOC	[n.d.]	
ADNM	ʃolo'vʲkɪ	

(b) the stem *ʲl(j)ud-*

	SG	PL
NOM		ʲludɪ
VOC		ʲludɪ!
ACC		ʲlu'dej
GEN		ʲlu'dej
DAT		ʲludʲam
INST		ʲludʲmɪ
LOC		ʲludʲax
ADNM		

(c) the stem *duf-*

	SG	PL
NOM		du'ʃɪ ¹²
VOC		
ACC		
GEN		duʃ
DAT		
INST		
LOC		
ADNM	duʃɪ	

The reader may have noticed that what I have included here (Tables 4 a, b, c) as the VOCATIVE (SINGULAR) form (*ʃolo'vʲʃe!*) can be either a derived form (with an appreciative), or a non-productive type of stem alternation. In favor of the second position, it must be said that the alternation between *-k* and *-ʃ* is almost certainly related to a phonological process from the Common Slavic era known as the “First Palatalization” (Shevelov 1964). This is certainly an old remnant of that, especially considering that the VOCATIVE is a marginal CASE value and perhaps more susceptible to retain older forms (see the discussion on CASES in: Corbett 2008, Daniel & Spencer 2009, Zaliznjak 2002). This also happens with some of the Belarusian and Russian remnants of the older VOCATIVE; e.g. (CSR) [NOM SG] *Bog* > [VOC SG] *Bože!* ‘oh, God!’.

4.2.1. Homophony and ambiguity

One of the biggest challenges for the analysis of the corpus has been the ambiguity caused by the homophony of the stems *ʃolo'vʲk-* and *duf-*. The latter derives from *dufa* ‘soul’ (most likely motivated by metonymy). The form *duf(a)* can be also found unbound, but it denotes literally ‘soul (i.e. spirit)’.

Corbett (2007) and Vanhove (2017) describe a similar problem with a few CSR suppletive nouns such as *reběnok-deti* ‘child’. The noun *reběnok* is not available for the PLURAL; and so, for the rest of the forms, it uses the stem *det-*; i.e. [NOM PL] *deti*; but not **reběнки*. The noun *ditjo* ‘child’ exists on its own, but is rather archaic and restricted to the literature. As a result, the most ‘direct’ or semantically regular [NOM PL] of *reběnok* is *deti*.¹³

In short, the fact that *dufa* can stand on its own and has a meaning closely related to ‘person’, does not impair part of its paradigm from being used as a suppletive form of ‘person’ (rather than it being a mere synonym). Moreover, we

¹² I have only recorded one instance of *duf-* being used (unbound) in the NOM PL, where it seems to mean ‘people’ instead of ‘souls’ (17).

¹³ Some people may argue that this noun also has a NOMINATIVE PLURAL and VOCATIVE PLURAL form, based on truncation [NOM PL] *rebjata!*; [VOC PL] *rebjat!* Although these two forms may be originally related to the SINGULAR form *reběnok*, the correlation between the two is not semantically regular. The SINGULAR forms ought to be translated as ‘child’; whereas the PLURAL forms are an informal way of addressing teenagers or adults, and so they could be translated as ‘guys’ (see also, Daniel & Spencer 2009).

have the cross-Slavic survey (Table 1, Table 2 and Table 3) as evidence of this form being employed as a suppletive form of the ADNUMERATIVE forms, most remarkably in Bulgarian (Table 1). Consequently, from now on, whenever there is a possibility of confusion, I will refer to *dufa*¹ as one of the possible suppletive stems of ‘person’; and *dufa*², as a proper non-defective noun, meaning ‘soul’.

The forms of *ʃolovik*- are far more complex. There is a continuous overlap of forms. Nevertheless, the noun *ʃolovik*¹ ‘person’ is, at least, homophonous with *ʃolovik*² ‘man, husband’, which I am also going to distinguish with supra-indexes for the sake of clarity.

(6) (B21.6 00:43)

ʃolovik ²	u	jijɪ	umer	u	sorok	pjat
husband.NOM.SG	in	3SG.GEN.F	die.PRF.PST.M.SG	in	forty	five.NOM
fod	u	babɪ	ot	tifu		
year.GRADNM	in	lady.GEN.SG	from	typhus.IIGEN.SG		
‘This lady’s husband (lit. man) died from typhus at the age of forty-five’.						

The evidence for this distinction is that in some varieties these nouns inflect differently in the cell of the ADNUMERATIVE. When it means ‘man or husband’ it has a prototypical ADNUMERATIVE and GEN PL/GRADNM of inflectional class II, and syntactically it behaves as a PLURAL noun, as we can see in the examples (7) – (10) (note that (8) and (9) are part of the same text).

(7) (T11.6 00:25)

tut	dva	ʃolo ¹ vɪkɪ ,	dva	‘xloptsɪ
here	two.NOM.M	man.ADNM	two.NOM.M	boy.ADNM
zabɪ-l-o		mol’niʃa		
kill.PRF-PST-N.SG		lightning(N).NOM.SG		
‘The lightning killed two men, two boys here’.				

(8) (T8.5 02:38)

muʒi ¹ k-iv	vs-ix ,	ʃolovi ¹ k-iv ²	ostavɪ-l-ɪ	na	nɪʃ,
male.GEN.PL	all-GEN.PL	man-GEN.PL	leave-PST-PL	in	night.ACC.SG?
noʃova-l-ɪ	vonɪ	tam			
overnight-PST-PL	3PL.NOM	there			
‘All the males, the men were kept [there] for the night, they spent the night there’.					

(9) (T8.5 05:03)

nas	nɪ	pobɪ-l-ɪ,	‘lud-aj,	nɪk-oʃo.	ɪ	ʃolovi ¹ k-iv ²
1PL.ACC	NEG	beat.PRF-PST-PL	person-ACC.PL	nobody-ACC.SG	and	man-ACC.PL
tɪx,		muʒi ¹ k-iv	vipusti-l-ɪ	ʃferez	nɪʃ	
that-ACC.PL		male-ACC.PL	release.PRF-PST-PL	through	night.ACC.SG	
‘And they did not kill any of our people; and those men, those males were released after the night’.						

(10) (TOR1.25 06:51)

tri	uzrozɪ-ix	ʃolo ¹ vɪkɪ	xodɪ-l-ɪ
three.NOM	adult-GEN.PL	man.ADNM	go.IMPV-PST-PL
‘Three adult men were walking’.			

I reiterate that such a distinction is not clear-cut and that for many varieties both lexemes inflect the same way.

4.3. Conditions for suppletion for the noun ‘person’

Suppletion makes inflectional paradigms more complex; and thus, less predictable and demanding more memorization effort (Matthews 1991). Ideally, we would expect the speakers to adhere to one (or maximum two) of the aforementioned paradigms (Tables 4 a, b, c), but I show how this expectation is far from being met.

In the previous sub-sections, particularly in §4.1, I have demonstrated that the stems involved in this suppletion riddle, were true suppletive stems and not simply synonyms. Now, we could state that there is pure overabundance (Thornton 2011, 2013, 2019), i.e. that two (or more) inflectional forms share the same cell and that they can be used interchangeably in any context (e.g. English *burned/burnt*). However, this is traditionally seen as a result of an inaccurate or superficial analysis by many linguists. In order to argue for this, I have decided to test the stems in different syntactic contexts or conditions to see if they restrict the use of certain stems. The putative conditions are based on the ones set by Bortnik (1979),¹⁴ for the Russian cognates of the suppletive nouns ‘year’ and ‘person’. In addition to these, I propose additional putative conditions based on observations of the behavior of numeral phrases (including quantifiers and various types of numerals)¹⁵ in the entire corpus of West Polesian.

4.3.1. Unbound or non-governed ‘person’

Even if most of the instances of ‘person’ analyzed here are related to quantification, let us start by looking at contexts in which the noun does not appear governed by a quantifier or NumP. The West Polesian corpus survey shows that all the three stems under discussion are possible even though the suppletion patterns vary from one speaker to another. In any case, as far as data in the corpus can attest, the noun ‘person’ uses at least two different stems in every speaker’s variety.¹⁶

- (11) (Z1.6.1. 03:51)
 dofian¹a-je nas **ʃfolo**‘vik^{2?}
 get-3S.SG 1PL.ACC man.NOM.SG
 ‘The man gets us’.
- (12) (Z1.6.1. 05: 33)
 ʃto **ʔudiam** bud-e, te jej nam!
 what.NOM people-DAT.PL BE.FUT-3SG same 1PL.DAT
 ‘Whatever is to happen to the rest of the people, shall it also happen to us!’

¹⁴ In Chumakina et al. 2004: 294)

¹⁵ I tried adding an extra condition: ‘the noun ‘person’ with collective numerals’; but, unfortunately, there are no instances of collective numerals + ‘person’ in the corpus.

¹⁶ I.e. it displays a suppletive paradigm in the speech of every participant for whom there is enough data, unlike the suppletive noun ‘year’.

As for the unbound forms, the stem *duf-* is certainly dispreferred, if not ungrammatical. This suggests that, most likely, it replicates the behavior of its cognate in Bulgarian and Macedonian. It is possible to find the form *dufa*² unbound, but as I have said, it literally denotes ‘soul; spirit’. Not surprisingly, it seems impossible to say something like ‘nasty people’ using *duf*⁻¹ (**biztolkovı dufı*?), but I have found examples of the other two in the corpus. See examples (13) – (15):

- (13) (TOR1.29 00:35)
 xoroʃ-ıje t-ıje **’ljudı**
 good-NOM.PL DEM-NOM.PL person.NOM.PL
 ‘Those good people.’
- (14) (TL6.3 01:59)
 pri-ʃ-ov star-enjk-ı **ʃolo’vık**
 arrive.PRF-PST-M.SG old- APPRECIATIVE -M.SG man/person.NOM.SG
 ‘An old man arrived.’
- (15) (Tor 1.12 02:11)
 tak-ıja **dobr-ıja** **’ljud-k-ı**
 DEM-NOM.PL good-NOM.PL person-APPRECIATIVE-NOM.PL
 ‘Such nice (little) people.’

In any case, I have only found two instances of unbound *dufa*¹ in the corpus (16), (17), and both of them doubtful.

- (16) (B1.1 19:20?)
 odno-jej **duʃı** ne pusti-v
 one-GEN.SG.F person/soul?GEN.SG NEG let-PST.SG
 ‘[He] wouldn’t let a single person in [lit. ‘not a soul’]’.
- (17) (TL1.1 16:36)
 nıeskolko ʃıo’dov sobıra-l-ı-sı, ı **du’ʃı** [nu]
 some year.GEN.PL gather.IPFV-PST-PL-REFL and people.NOM.PL well
 sosıedıe v asnobnom, vsıe prixodı-l-ı
 neighbor-NOM.PL in majority all-NOM.PL come.IPFV-PST-PL
 ‘[We] were gathering for several years, and people, that is to say, mostly neighbors, all used to come’.

In the first example (16), the stem *duf-* is not governed by a lower or higher numeral, the preceding constituent is a quasi-adjectival numeral (see Corbett 1978). The second example (17) comes from another elderly speaker, whom I have recorded using the stem *duf-* with numerals as a suppletive form of ‘person’. However, it may be that the speaker really had in mind its primary meaning ‘soul’ when she produced this sentence.

4.3.2. Bound forms

Most of the conditions which could have an effect on the distribution of the stems are those syntactic contexts in which ‘person’ is governed.

4.3.2.1. Bound to a cardinal numeral

When the noun ‘person’ is governed by a cardinal numeral all three stems can appear.

- (18) (TOR1.25 07:50)
 t-IX Xavansk-IX tože **adnatsat** **duf**,
 that-ACC.PL Khavansk-ACC.PL also eleven person.GRADNM
ʃilo'viek zavizli¹⁷
 person.GRADNM bring.PRF-PST-PL
 ‘They also brought eleven people from Khavansk’.
- (19) (TL6.2 00:48)
 [...] **dvatset** **ʃolo'vik** v komnat-i
 twenty person.GRADNM in room-LOC.SG
 ‘[...] twenty people in the room’.
- (20) (TOR1.47 01:47)
 tam **miljon-i** **lu'dej** zakopan-ix
 there million-ACC.PL person.GEN.PL buried-GEN.PL
 ‘There are millions of people buried there’.

4.3.2.2. With question words (Q) and quantifiers

When a question word (Q) or a quantifier is governing the NP, I have documented the stems *ʃolovik-* and *lud-*, and so far, I have not found the stem *duf-* used for this. Nevertheless, I do not see strong reasons why it could not be employed in these contexts as well.

- (21) (TOR1.47 04:20)
 z dirjevni vizva-l-i **mnio** **lu'dej**
 from village.GEN.SG call.PRF-PST.PL many person.GEN.PL
 ‘They called out many people from the village’.
- (22) (TL3.2 00:44)
nieskolko **ʃolo'vik** povjerova-l-i
 some.NOM person.GRADNM believe.PRF-PST-PL
 ‘Some people came to faith’.
- (23) (B19.3.0 00:26)
 skaz-ete mn'i kofo vi baʃi-l-i, **kilka**
 tell.IMP-2PL 1SG.DAT Q.ACC 2PL.NOM see-PST-PL how_many
ʃolo'vik vi baʃi-l-i, xto bud-e, xto **bilf**
 person.GRADNM 2PL.NOM see-PST-PL REL.NOM BE.FUT-3SG REL.NOM more
 ufilied-it **lu'd-ej** tomu ja spodnits-u kup-lju
 see.PRF-3SG person-GEN-PL REL-DAT.SG 1SG.NOM skirt-ACC.SG buy.PRF-1SG
 ‘Tell me how many people (men) you saw. Whoever sees more people, I will buy her a skirt’.

¹⁷ Note that the speaker was aware of the suppletive stem *duf-*, but probably in an attempt to be understood by me, an outsider, she tried to simplify the paradigm for me.

4.3.2.3. With the adjective ‘all’

With the adjective *vsī* ‘all’ only the stem *lud-* has been attested, and it seems unlikely that *ʃolovik-* or *duf-* would be allowed in such a context.

- (24) (HL2.16 02:21) [HL4 speaking]
 [...] vs-ix lu'd-ej v adn-u kup-u [sic]
 all-ACC.PL person-ACC.PL in one-ACC.SG pile-ACC.SG
 ‘[...] all the people in one pile.’

4.3.2.4. Before the numeral

In West Polesian when a cardinal numeral is used after the noun it determines, it indicates that the quantity is approximate (rather than exact). Such a pattern is not only commonly shared by other Slavonic languages, but according to Greenberg’s 44th Generalization (1978: 284), it is a frequent cross-linguistic phenomenon. In the examples **Ошибка! Источник ссылки не найден.** and (26), both produced by B18 within the same story, we can see that in **Ошибка! Источник ссылки не найден.**, the main character had collected exactly two kilos [of berries] (B18 referred to a specific amount), whilst in (26) the character is negotiating with someone who looks at his bucket of berries and calculates an approximate quantity.

- (25) (B18.4E 00:37)
 [...] dva 'kil
 two kilo.ADNM
 ‘Two kilos’.
- (26) (B18.4 00:33)
 nu, ʃtʃo ʃ, Xonj-e, kil dva bud-e
 so what PART Xonja-VOC kilo.GEN.PL/GRADNM two.NOM.N be.FUT-3SG
 ‘So, what, Xonja, there’ll be around two kilos’.

In these instances the NP stands in GEN PL (or GR. ADNUMERATIVE), although the numeral may be a LOWER one. Concerning suppletion, so far, I have only documented *ʃolovik-* and *duf-*, but not *lud-*. This restriction is also shared by Russian, so there seems to be evidence to believe that the stem *lud-* is ungrammatical in this type of construction.

- (27) (B20.8 00:11)
 zbra-l-a-sa kompanja; ʃolo'vik deset, dvanatsat
 gather-PST-F.SG-REFL company.NOM.SG person.GRADNM ten twenty
 pri... primjerno odn-oho ʃod-u.
 for_example one-GEN.SG.M year-II-GEN.SG
 ‘A group used to gather, about ten people or twelve people of more or less the same age (lit. year)’.

- (28) (T1.3 03:18)

ʃtʲirɪ	stola,	ʃolo'vɪk	ʃestdisʲat,	vosimdisʲat
four.NOM	table.GEN.SG	person.GRADNM	sixty	eighty
i	ʃtʲirɪ	dnʲ-a.		
and	four.NOM	day-GEN.SG		

‘Four tables, around seventy or eighty people and four days’.

(29) (TL1.1 16:45)

duf	diesʲet	navjerno, [...]	ja	tak	pridpolofia-ju
person.GRADNM	ten	probably	1SG.NOM	so	guess-1SG
duf	diesʲet	pokajʲi-l-o-sʲa			
person.GRADNM	ten	repent-PST-N.SG-REFL			

‘About ten people, most likely <...> I guess that around ten people repented’.

4.3.2.5. Approximate quantity

When an approximate quantity (rather than an accurate cardinal form) is used, the stem *lud-* is not permitted.

- (30) (Z4.1.1 09:04)

xo'dɪ-l-i	po	tri,	po	pʲat	ʃolovɪk
go.IPFV-PST-PL	in	three	in	five	person.GRADNM
i	ʃ-l-i	ʊ	lis		
and	go-PST-PL	to	wood.ACC.SG		

‘[They] used to go to the forest [in groups of] three -five people’.

(31) (T1.1 24:03)

tam	bu-ʊ	ʃolovɪk	moʒe	tɪstʲʃa
there	BE-PST.M.SG	person.GRADNM	maybe	thousand

‘There were about three hundred people.’

(32) (T1.18 01:16)

pʲat,	siem	duf,	na	siem	ʃolo'vɪk
five	seven	person.GRADNM	for	seven	person.GRADNM

‘Five or seven people, [enough] for seven people’.

4.3.2.6. Perception of quantity

I have found instances of the stems *ʃolovɪk-* and *lud-*, with larger and smaller numbers, but not *duf-*, probably because there are not many instances of the stem *duf-* in the corpus. However, there is no evidence that the choice of the stems is related to any pragmatic/semantic factors; i.e. whether the speaker perceives the number of people as small (and thus, easily individuated) or large. In (33) -(35) I show some examples of all the stems being used with large amounts.

- (33) (T11.2 00:31)

bu-l-o	po	sto pʲitdisʲat,	po	dvjestɪ
BE-PST-N.SG	about	hundred_and_fifty	about	two_hundred
ʃolo'vɪk	na	svadv-e	ʃulʲa-l-o	kolʲisʲ...
person.GRADNM	in	wedding-LOC.SG	party-PST-N.SG	formerly

‘About 150 -200 people used to party in [each] wedding’.

- (34) (TOR1.6 00:50)
 vin **tɨsaʈʃɪ** **lu'dej** [...] pap-a spas
 3SG.NOM.M thousand person.GEN.PL father-NOM.SG save.PRF.PST.M.SG
 'It can be said] that [my] father saved a thousand people'.
- (35) (TL1.1 17:43)
dʲevjenosto dvje **du'ʃɪ** **bu-l-o** v dom-i
 ninety-two.NOM.F person.ADNM BE-PST-N.SG in house-LOC.SG
 'Ninety-two people were at home'.

4.3.3. Sociolinguistic remarks

The West Polesian speech community lives in a marshy area which has been very difficult to access until recently, keeping it isolated for centuries. Seeing as West Polesian is a non-standardized and minoritized variety suddenly exposed to four standardized Slavic varieties (Polish, CSR, BLM and ULM), language change is accelerating and it is not uncommon to find loans (both lexical and grammatical) from its neighbors. One may be tempted to argue that language contact is responsible for all the irregularities in the paradigms. However, in this case, this is not an entirely plausible answer. As I have shown in §4.1, none of the four surrounding Slavic varieties have the stem *duf-* in their inventories (synchronically).¹⁸ The fact that the stem is present in other Slavic varieties (e.g. Bulgarian), with which the community has not had contact for centuries confirms that this is rather a Common-Slavic peculiarity that most contemporary West and East Slavic varieties have lost.

Having said this, I have only been able to document the stem *dufa* (when it is used as a suppletive form of the noun 'person') in the speech of the older generation (born before 1950): B1, T1, TL1, Z4 and Z10, who also happen to be all female, except for T1. Nevertheless, my intuition is that the fact that only women appear to use this form in my corpus is either accidental or motivated by an unbalanced gender sample,¹⁹ especially concerning the oldest speakers. Hence, the alternation between *duf-* and the other two stems seems more strongly conditioned by age. I should remark that age is often strongly related to bilingualism/exposure to other Slavic varieties, particularly Russian. On the one hand, children (particularly women) who were of school age during World War II often missed several grades

¹⁸ It was used in Russian and Polish up to the end of the 19th century or the beginning of the October Revolution to refer mainly to servants (see for example Gogol's *Dead Souls*), i.e. the use was more nuanced and stylistically marked.

¹⁹ I admit there is a gender imbalance in the corpus. From the whole corpus, only fifteen men (27.77%) took part at least in some way, in contrast with thirty-nine women (72.22%). Besides, men's interventions were considerably shorter than women's (often participating as 'people in the back'). This is related to two main factors. First, life expectancy for men is considerably shorter than for women in the area; in fact, a large number of the women I interviewed were widows. Second, in the rural areas where I carried out fieldwork most men under fifty-five work jobs outside of their village (primarily seasonal work in the building industry). Thus, finding men for interviews was more challenging.

(at the time, most received only four years of schooling, but many women only completed one or two). Moreover, people who attended school before World War II (under Polish rule) received their education in Polish (and thus often can be illiterate in Russian and Belarusian).²⁰ Younger people tend to have received more schooling, which since WWII has been conducted in CSR or BLM. Moreover, younger speakers have generally spent or still spend more time in the cities and towns where Russian is the primary language. Hence, the younger the speaker, the most likely they are to be bilingual in Russian, and hence more likely to code-switch.

I have not been able to find any instances of *lud-* in TL1's recordings. This could make us suspect that she does not use the stem at all, although it seems very unlikely. I used to have a similar situation with Z4, who used *dufa* (as a suppletive root) in one of the sessions, but I have eventually been able to document a token of *lud-*, in an OBLIQUE CASE (37). Compare (36) to (37) and both produced by Z4.

- (36) (Z4.1.2 00:49)
 z nas fest duf: bu-l-o dtej festero [...]
 from 1SG.GEN six person.GRADNM BE-PST-N.SG child.GEN.PL six(COLL)
 'Altogether, we were six people: there were six children [together] <...>'.
 (37) (Z4.1.1 03:41)
 [...] itı po 'ludl-ax, moze tam komu ffo pomofi-tı
 go-INF for person-LOC.PL maybe there Q.DAT Q.ACC help-INF
 '<...> going to different people, maybe someone could help'.

Thus, we can at least affirm that, although a minority, the three suppletive stems are present in some idiolects. And by looking at their behavior in other Slavic languages, we have more evidence to state that they are not overlapping synonyms, admitting that some stems are more frequent than others.

Finally, considering other sociological and/or sociolinguistic factors such as class/economic status, the use of one form or the other is not marked by that. The reason for this is that all speakers interviewed belonged to the same social class (rural, peasants or blue-collar workers) and that speaking West Polesian is already quite stigmatized in Belarusian Polesie. Thus, people who try to 'pass' as a person belonging to a higher social class would most likely try to speak Russian, and less often, Standard Belarusian. This sort of sociolinguistic setting is not uncommon in many speech communities in the world, although, perhaps, the first person to describe a similar situation thoroughly was Nancy Dorian (2010) and her work on Suffolk Gaelic.

5. Summary and conclusions

I have presented the problem of the suppletive noun 'person' in West Polesian. I have started by asking whether the alleged three stems belonged to the same

²⁰ One of the older speakers who was schooled during the Nazi occupation reported receiving her education in Standard Ukrainian. The remaining people interviewed were schooled in Polish, Russian or Belarusian, if they received any formal education.

paradigm or whether they are synonyms. Evidence from the cross-Slavic survey has shown that these stems also exist in other better documented Slavic varieties as part of their suppletive paradigms. Narrowing down the focus to the West Polesian corpus, I have then applied some putative conditions that would determine the use of one stem or the other. Once these conditions are applied, the resulting picture can be found in (Table 5), which is far from homogeneous. Having said this, there is a handful of patterns that can be extracted considering the putative conditions and the use of the stems by different speakers:

- The noun ‘person’ has a suppletive paradigm in every speaker’s idiolect, which is usually composed of two stems.
- The stem *ʃolovik-* appears in the NOM SG of every idiolect, but cannot appear in the NOM PL.
- The stem *lud-* is not allowed with approximate quantities nor in a prenumeral position.

Nevertheless, there is still a vast amount of variation regarding the distribution of the suppletive stems across speakers of the same village (cf. TL1 with TL3-TL4), and even within a single speaker’s idiolect and, hence, overabundance. For this reason, the speakers in (Table 5) are not arranged according to their origin or alphabetical order. Instead, I have decided to group them based on similarities in their suppletive patterns.

Table 5. Distribution of the suppletive stems based on utterances from the most recurrent speakers in the corpus

	NOM SG	LOWER NUMERALS	HIGHER NUMERALS	NOM PL	PRENUMERAL POSITION	+ QUANTIFIER OR Q WORD
Z10	ʃolovik ²		duʃ			
Z4			ʃolovik			
			duʃ			
TL1	ʃilovjek	duʃɪ		duʃɪ	duʃ	
T1	ʃolovik	ʃolovikɪ	ʃolovik	ludɪ	ʃolovik	ludej
		ʃolovika	duʃ			
B20	ʃolovik		ʃolovik	ʃudɪ		
TOR1	ʃolovik	ʃolo'vɪkɪ	ʃolovik	ludɪ	ʃolovik	ludej
			ludej			
T11		ʃolovikɪ	ʃolovik			ludej
TL3		ʃolovika	ʃjelovjek	ʃudɪ		ʃolo'vɪk
			ʃolovik			
TL4	ʃolovik		ʃolovik	ludɪ		
				ʃudɪ		
HL2		ʃolovika		ludɪ		

The results from the study, rather than being very conclusive, open the door to advance in several bigger questions in morphosyntactic typology and variation studies. First, they challenge some of the assumptions concerning suppletion and the need for some regularity supported by some morphosemantic patterns for the sake of learnability (cf. Bybee 1985, Matthews 1991). Second, morphological overabundance has been often related to purely morphological causes. For example,

Thornton (2019) argues that the origin of morphological overabundance is to be found in either suppletion or heteroclisis. This also applies to the West Polesian noun ‘person’. Yet, there are some sociolinguistic factors that may have also motivated the situation at hand and which would be overlooked by most traditional theoretical frameworks. Finally, the findings from this paper also add evidence to Dorian (2010) and Thornton’s (2013) claims that sociologically neutral inter- and intra-speaker variation and morphological overabundance are far more common than most traditional linguistic descriptions and theories want to admit. And for that matter, the role of usage-based corpora from understudied languages will increasingly become relevant in this new theoretical and descriptive wave.

Abbreviations

Glossing abbreviations in this paper are based on the Leipzig Glossing Rules (2015 revision), with some additions:

1, 2, 3: first, second, third person (respectively)

ACC: accusative

ADNM: adnumerative

BG: Bulgarian

BLM: Standard (Literary) Belarusian

COLL: collective numeral

CSR: Contemporary Standard Russian

CZ: Czech

DAT: dative

DEM: demonstrative

DET: determiner

F: feminine

FUT: future

GEN: genitive

IIGEN: second genitive

GRADNM: greater adnumerative

HR: Croatian

IMP: imperative

INS: instrumental

IPFV: imperfective

LOC: locative

M: masculine

MKD: Macedonian

N: neuter

NEG: negation

NOM: nominative

NP: noun phrase

NumP: numeral phrase

PART: particle

PL: plural

POL: Polish

POSS: possessive

PRF: perfective
PRS: present
PST: past
Q: question particle/marker
REFL: reflexive
REL: relative
SG: singular
SK: Slovak
SVO: Slovene
ULM: Standard (Literary) Ukrainian
US: Upper Sorbian
VOC: vocative
WP: West Polesian

Acknowledgements

First of all, I would like to thank a thousand times my dear *Poleshchuki*, who have contributed in different way to this project: either as language-consultants or chefs, drivers, sauna-masters, conversation partner. Thank you for your patience and incredible hospitality! Fieldwork research, where the data here presented were obtained was possible thanks to the generous support of the Belarusian Charity Trust Ltd, Gilchrist Education Trust and the Philological Society (UK). The ELDP grant IPF-0189 also granted the space and money to synthesise this work. I hereby declare no conflict of interest and I assume all responsibility for any mistakes. I am grateful for the support and suggestions of Greville Corbett, Matthew Baerman and Oliver Bond during the entire research, as well as the comments from other SMG members and visitors, particularly Anja Hasse, Hellen SimsWilliams and Anna Thornton. And last, but not least, many thanks to Rob Tegethoff for revising this article and to Diana Forker for encouraging me not to drop academia.

REFERENCES

- Akiner, Shirin. 1983. The Syntax of the numeral in Byelorussian, compared with Ukrainian, Russian and Polish. *The Slavonic and East European Review* 61 (1). 55–68.
- Babby, Leonard H. 1987. Case, prequantifiers, and discontinuous agreement in Russian. *Natural Language & Linguistic Theory* 5 (1). 91–138. <https://doi.org/10.1007/BF00161869>
- Bortnik, Ninel'. 1979. Nekotorye osobennosti osnov obrazuyushchikh dvoynye formy mnozhestvennogo chisla. *Russian Language Journal* 32. 43–58.
- Bybee, Joan L. 1985. *Morphology. A Study of the Relation Between Meaning and Form*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Chumakina, Marina, Hippisley Andrew & Corbett Greville G. 2004. Istoricheskie izmeneniya v russkom yazyke: Sluchai chereduyushchegosya suppletivizma. *Russian Linguistics* 28. 281–315.
- Corbett, Greville G. 1978. Numerous squishes and squishy numerals in Slavonic. In B. Comrie, (ed.), *Classification of grammatical categories*, 43–73. Edmonton, Canada: Linguistic research INC.
- Corbett, Greville G. 1983. *Hierarchies, Targets and Controllers: Agreement Patterns in Slavic*. London: Croom Helm.

- Corbett, Greville G. 2005. The canonical approach to typology. In Z. Frajzyngier, A. Hodges & D. S. Rood (eds.), *Linguistic diversity and language theories*, 25–49. Amsterdam: John Benjamins.
- Corbett, Greville G. 2007. Canonical typology, suppletion, and possible words, *Language* 83 (1). 8–42.
- Corbett, Greville G. 2008. Determining morphosyntactic feature values: The case of case. In G. Corbett & M. Noonan (eds.), *Case and grammatical relations: Papers in honour of Bernard Comrie*, 1–34. Amsterdam: John Benjamins.
- Daniel, M. & Spencer A. 2009. The vocative-an outlier case. In A. Malchukov, A. Spencer (eds.), *The Oxford handbook of case*, 626–634. Oxford: Oxford University Press.
- Dorian, Nancy C. 2010. *Investigating Variation. The Effects of Social Organization and Social Setting*. New York/Oxford: Oxford University Press.
- Franks, Steven. 1995. *Parameters of Slavic Morphosyntax*. Oxford University Press.
- Greenberg, Joseph H. 1978. Generalizations about numeral systems. In J. H. Greenberg et al. (eds.), *Universals of human language*, 249–295. Stanford (CA): Stanford UP.
- Kim, Hyoungsup. 2009. *The Sstructure and Use of Collective Numeral Phrases in Slavic: Russian, Bosnian/Croatian/Serbian, and Polish* [PhD Thesis]. The University of Texas at Austin.
- Madariaga, Nerea & Igartua Iván. 2017. Idiosyncratic (dis)agreement patterns: The Structure and diachrony of Russian paucal subjects. *Scando-Slavica* 63 (2). 99–132. <https://doi.org/10.1080/00806765.2017.1390922>
- Matthews, P. H. 1991. *Morphology*. 2nd ed. Cambridge: Cambridge University Press.
- Mel'čuk, Igor A. 1994. Suppletion: Toward a logical analysis of the concept. *Studies in Language* 18 (2). 339–410.
- Neset, Tore. 2019. Big data in Russian linguistics? Another look at paucal constructions. *Zeitschrift für Slawistik* 64 (2). 157–174. <https://doi.org/10.1515/slav-2019-0012>
- Neset, Tore & Maria Nordrum. 2019. Do Russian paucal numerals govern the genitive? Evidence from stress placement. *Russian Linguistics* 43 (2). 87–105. <https://doi.org/10.1007/s11185-019-09209-7>
- Nurmio, Silva & Willis David. 2016. The rise and fall of a minor category: The case of the Welsh numerative. *Journal of Historical Linguistics* 6 (2). 297–339. <https://doi.org/10.1075/jhl.6.2.05nur>
- Pereltsvaig, Asya. 2013. On number and numberlessness in languages without articles. In C. Cathcart et al. (eds.), *Proceedings of the 37th annual meeting of the Berkeley linguistics society*, 300–314.
- Roncero, Kristian. 2021. Revisiting numeral phrases in East Slavic: Insights from West Polesian. *Slavonic and East European Review* 99 (4). 601–646.
- Shevelov, George Y. 1964. *A Prehistory of Slavic. Historical Phonology of Common Slavic*. Heidelberg: Winter Univ. Verl.
- Thornton, Anna M. 2011. Overabundance (multiple forms realising the same cell): A non-canonical phenomenon in Italian verb morphology. In M. Maiden et al. (eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology*, 357–381. Oxford: Oxford University Press.
- Thornton, Anna M. 2013. La non canonicità del tipo It. braccio // braccia / bracci: Sovrabbondanza, difettività o iperdifferenziazione? *Studi Di Grammatica Italiana* 29–30. 419–477.
- Thornton, Anna M. 2019. Overabundance: A canonical typology. In F. Rainer et al. (eds.), *Competition in inflection and word-formation*, 189–224. Dordrecht: Springer.
- Vanhove, Claudia. 2017. *A Diachronic View on Suppletion. Diachronical Research Regarding The Lexical Competition between Rebēnok and Ditja* [PhD Thesis]. University of Gent.

- Viellard, Stéphane. 2011. Problèmes de syntaxe des numéraux dans quelques langues slaves (vieux slave, vieux russe, BCMS, polonais, russe moderne), introduction. *Revue des Études Slaves* 82 (4). 597–603.
- Waldenfels, R. & Meyer R. 2011. *ParaSol, a Corpus of Slavic and Other Languages*. (Electronic resource): University of Bern/University of Regensburg. <http://parasol.unibe.ch>. (accessed 24 February 2022).
- Zaliznjak, Andrej A. 2002. *Russkoe Imennoe Slovoizmenenie: S Priloženiem Izbrannyx Rabot Po Sovremennomu Russkomu Yazyku I Obščemu Yazykoznaniju*. Moscow: Nauka.
- Žolobov, Oleg F. 2003. K istorii malogo kvantitativa: Adnumerativnye formy prilagatel'nyx i suščestvitel'nyx. *Russian Linguistics* 27. 117–197.

Other resources

- Hrvatski Jeyzčni Portal*. 2006–2021. Srce; Znanje. Available from: <http://hjp.znanje.hr> (accessed 3rd Apr 2021).
- Leipzig Glossing Rules*. 2015, May 31 update. Max Planck Institute for Evolutionary Anthropology. <https://www.eva.mpg.de/lingua/resources/glossing-rules.php> (accessed 24 February 2022).
- Oxford English Dictionary. 2018. *OED Online*. Oxford University Press. <http://www.oed.com> (accessed 24 February 2022).
- Rečnik na bālgarskija ezik*. 2011–2018. Institutāt za bālgarski ezik. <http://ibl.bas.bg/rbe/lang/bg/> (accessed 24 February 2022).

Supplementary materials

The supplementary materials (a list with all the tokens of ‘person’ in the West Polesian corpus) are available from <https://doi.org/10.5281/zenodo.5879004>

Appendix. Geographic varieties covered in this research

All the texts were collected in the region of Brest (Belarus).

Table 6. Varieties covered in the research with GPS coordinates

Village (local name in brackets), district	GPS Coordinates
Aliaksiejevičy, Drahičyn	52°17'03.5"N 24°59'34.2"E
Bahdanaŭka (Bodanyŭka), Luniniec	52°21'59.5"N 26°27'04.9"E
Haloŭčyci, Drahičyn	52°08'50.1"N 24°51'58.5"E
Imianin (Torokan), Drahičyn	52°14'09.6"N 24°54'26.9"E
Pare, Pinsk	51°53'49.7"N 26°08'10.8"E
Semekhavičy, Pinsk	51°56'54.5"N 25°52'13.5"E
Tatar'ja, Drahičyn	52°18'01.4"N 24°56'49.8"E
Tolkava, Drahičyn	52°11'33.0"N 24°54'48.9"E
Vostraŭ, Pinsk	51°53'49.7"N 26°08'10.8"E
Žyďča, Pinsk	51°57'26.3"N 25°56'39.3"E

Article history:

Received: 24 July 2021

Accepted: 14 December 2021

Bionote:

Kristian RONCERO is a postdoctoral researcher at the Max Planck Institute for Evolutionary Anthropology and the Friedrich Schiller University in Jena (Germany). His research focus is on Chamalal (Nakh-Dagestanian); West Polesian (East Slavic); morphological typology; and language description. He is the author of an online corpus: *An audio-visual documentation of Chamalal, a language of Dagestan* (2019-2021, ELAR) and is due to join soon the *Feast and Famine* project (University of Sheffield, UK) on morphological overabundance and defectivity.

Contact information:

Max Planck Institute for Evolutionary Anthropology

Deutscher Platz 6, 04103 Leipzig, Germany

e-mail: kristianroncero@gmail.com

Сведения об авторе:

Кристиан РОНСЕРО – постдокторант Института эволюционной антропологии им. Макса Планка и Университета Фридриха Шиллера в Йене (Германия). В центре его исследовательского интереса – чамалинский (нахско-дагестанский) и западнополесский языки, морфологическая типология и описание языков. Он автор онлайн-корпуса «Аудиовизуальная документация чамалинского языка, Дагестан (2019–2021, ELAR)».

Контактная информация:

Max Planck Institute for Evolutionary Anthropology

Deutscher Platz 6, 04103 Leipzig, Germany

e-mail: kristianroncero@gmail.com